W. Kells
October, 1978

I. Introduction

Refinement of the idea¹ of stochastic cooling of stored particle beams has converged rapidly in recent months to a practical demonstration.² The swiftness of this advance has raised several unresolved questions pertinent to extrapolating parameters of the method to a broader range of machine designs. The central questions are:

- A. Can the powerful performance of the demonstrated² momentum cooling be duplicated in a practical Betatron cooling scheme?
- B. How does cooling performance (e.g., rate) depend on cooling beam intensity (or, more generally, warm beam phase space)?
- C. How can we identify performance limitations which are due to the statistical nature of the signal sample from those due to amplifier noise? For instance, is the cooling rate of the CERN accumulator design limited by amplifier quality or by beam intensity?

Another poorly understood concept is that of "mixing", that is the role of the time scale for coherence of a given beam segment. However, this question may only be properly understood in conjunction with the others. It cannot be viewed independently.

The object of this paper is to answer as much about the above questions as possible in as clear a manner as possible.

II. Basic Betatron Cooling

Consider a beam of one particle. Its betatron amplitude is A

and it has phase ψ at the P.U. position (Fig. 1). If the electronic channel is truly symmetric with respect to the particle orbit, there is a correction:

$$\begin{aligned} A^2 &\rightarrow A^2 + g^2 (A \cos \psi)^2 - 2g[A \sin(\psi + \frac{\pi}{2})] [A \cos \psi] \\ &= A^2 - A^2 \cos^2 \psi [2g - g^2] \end{aligned} \quad (1)$$

where we take this equation as a definition of g , which has lumped in it factors for P.U. & K sensitivity, electronic bandwidth, and amplifier gain. Electronic noise is ignored at this stage.

To introduce N additional particles random in phase and time with respect to this particular one, we shall have to suitably average over their relative P.U. arrival times and phases. Assuming the electronic system to be linear, the kick will just be proportional to a sum of individual signals, as used in (1), over N :

$$S = \sum_{i=1}^N A_i f(t_i) \cos \psi_i + A \cos \psi \quad (2)$$

where, to be consistent with (1), $f(0) = 1$. $f(t)$ is a positive function describing the P.U. effective length (Fig. 2). Then equation (1) becomes,

$$A^2 \rightarrow A^2 + g^2 S^2 - 2gA \sin(\psi + \frac{\pi}{2} + \delta) S \quad (3)$$

Notice the δ representing local betatron wavelength spread (however I ignore this spread's effect on g , in the spirit of g being a mean lumped constant). Since the t_i and ψ_i are independent random variables we average independently. The averaging is faithfully

represented by the substitutions:

$$\begin{aligned} A_i^2 &\rightarrow \overline{A^2} \\ \cos^2 \psi_i &\rightarrow \frac{1}{2} \\ \cos \psi_i \cos \psi_j &\rightarrow 0 \quad i \neq j \end{aligned}$$

Then (3) becomes:

$$A^2 \rightarrow A^2 + \frac{1}{2} g^2 \overline{A^2} \left(\sum_{i=1}^N f^2(t_i) \right) - 2gA^2 \cos^2 \psi + 2\delta g A^2 \sin \psi \cos \psi \quad (4)$$

But A is just one of the A_i so:

$$\overline{\Delta A^2} = \frac{1}{2} g^2 \overline{A^2} \left(\sum_{i=1}^N f^2(t_i) \right) - g \overline{A^2} + \frac{2}{N} \delta g \overline{A^2} \sum_{i=1}^N \sin \psi_i \cos \psi_i \quad (5)$$

The last term also has zero average, meaning that, in general, spread in wavelength has no first order effect. For the first term a sensible definition is:

$$\sum_{i=1}^N f^2(t_i) \equiv N_s f^2(0) = N_s \quad (6)$$

where s stands for "sample". N_s is the average effective number of particles contributing to the P.U. signal at a given instant. Since the t_i are statistically independent, this definition is consistent with the intuitive notion that N_s scale linearly with N and with the P.U. length. We have, finally,

$$\frac{\overline{\Delta A^2}}{\overline{A^2}} = -\left(g - \frac{1}{2} N_s g^2\right) \quad (7)$$

The value of (7) is maximized with $g = N_s^{-1}$. This value of g is also precisely the one which corrects the signal S to zero at all points downstream from the kicker which are half integer betatron wavelengths. Therefore, no signal to be corrected would be available on subsequent passes through the P.U. (assuming an impossible machine tune of $m/2$, which will be shown unimportant to the cooling result). To consider the cyclical process, some mixing of the betatron phases must be assumed. There seems to be no way to accomplish this within a fixed sample so that we must have longitudinal mixing of (statistically independent) contiguous samples during each period. The analysis leading to (7) assumes no mixing between P.U. and K., so we are faced with a somewhat contradictory requirement. However, assuming ideal mixing, and observing that noise can only make matters worse, we obtain a $1/e$ shrinking time for $(\overline{A^2})^{1/2}$ of,

$$\tau = \frac{4LN}{\beta c} = 4N_s T_o$$

where L = effective P. U. length (8)

T_o = mean period

$$= 4 \cdot s \quad (N = 10^9; L = 30 \text{ cm}; \beta \approx 1)$$

Now I will upgrade (7) to include electronic noise. Defined in terms of an equivalent beam displacement, noise (at $t = 0$ in the sense of (2)) contributes a new term to (2):

$$S = \sum_{i=1}^n A_i \cos \psi_i + r \quad (9)$$

In following through once again to (7) only the r^2 term survives the averaging, leaving:

$$\Delta \overline{A^2} = g^2 r^2 - (g - \frac{1}{2} N_s g^2) \overline{A^2} \quad (10)$$

The solution of this equation is:

$$\overline{A^2}(t) = \overline{A^2}(\infty) + (\overline{A^2}(0) - \overline{A^2}(\infty)) e^{-2t/\tau} \quad (11)$$

$$\text{with } \overline{A^2}(\infty) \equiv \frac{2gr^2}{2-N_s g}$$

That is, noise determines only the asymptotic beam width. For the example (8) we may ask what value $\overline{A^2}(\infty)$ will have for given r and the maximal g .

$$\overline{A^2}(\infty) = 2 \frac{r^2}{N_s} \quad (12)$$

Actually (9) is a simplification since it implies both signal and noise source are at the same circuit point. The inverse N_s dependence is, however, correct, and to go further I refer to the appendix for a more precise result (Fig. 3 and Eq. A3).

$$\overline{A^2}(\infty) = \frac{2d^2 C_o^2}{e^{2N_s}} KT(B.W.) (50\Omega) \quad (13)$$

But C_o and (B.W.) (Appendix) as well as N_s (Eq. 8) may be related simply to L giving:

$$\begin{aligned} \frac{\overline{A^2}(\infty)}{\overline{A^2}(0)} &= \left[\frac{d^2}{\overline{A^2}(0)} \right] \left[\frac{T_o}{N} \right] \left[\frac{4K}{e^2} \right] [50\Omega] [E_o C]^2 T \\ &= 16 \cdot \frac{10^{-6}}{10^9} s \cdot (1.2 \times 10^{17} \text{ } ^\circ K^{-1}) \cdot (377\Omega)^{-2} \cdot 200^\circ K = 12.0 \end{aligned} \quad (14)$$

where $(\overline{A^2}(o))^{\frac{1}{2}} = d/4$, and the 200°K temperature corresponds to 2db noise figure amplifier. The quantity of interest in transverse cooling is $\overline{A^2}(t)$ since emittance is proportional to it. It is obvious that very low electronic channel temperatures are needed to achieve significant emittance reduction.

III. Basic Comparison with Momentum Cooling

One way to cool in momentum would be as illustrated in Fig. 4. Since all particles produce identical P.U. pulses, the cooling signal must derive from P.U. - K. time of flight. Some differentiation is necessary to provide the correct symmetry. Linearizing the kick gives, for a single particle beam

$$P^2 \rightarrow (P - gP)^2 \quad (15)$$

where $P \equiv (p - p_0)/p_0$ is analagous to the betatron amplitude of Eq. 1. In the sense that no additional phase degree of freedom is relevant, momentum cooling is simpler than betatron and the notion of "mixing" does not apply. However, since dispersion is necessary, a longitudinal mixing is inevitable. The dispersion must be matched to the P.U./K. B.W. so that all particles "see" their coherent kick (justifying the linear approximation).

Adding N particles requires only a time average (once again times are relative to the first "test" particle):

$$S \equiv \sum_{i=1}^N f(t_i - \eta\lambda P) + f(0 - \eta\lambda P)$$

where (16)

$$\eta = (df/f)/(dp/p_0)$$

$$\lambda = \text{P.U.} \rightarrow \text{K. time of flight.}$$

To be consistent with (15) $f(\eta\lambda P) = P$, and the correction now is:

$$P^2 \rightarrow (P - gS)^2 \quad (17)$$

which is,

$$\overline{\Delta P^2} = -2g\overline{P^2} + g^2 \overline{\sum_{i=1}^N f^2(t_i - \eta\lambda P)}$$

The first term is identical (functionally) to the g term of Eq. 5. The g^2 terms are quite different, since $\sum_{i=1}^N f^2(t_i - \lambda \eta P)$ i.e., independent of P . Such a term must be treated as the electronic noise term was in Eq. 10:

$$\overline{P^2}(\infty) = gN_s \quad (18)$$

For a reasonable requirement $P(\infty) = 10^{-3}$, $g \leq 10^{-6}/N_s$, which is much too slow for interest. Some mechanism is needed to damp the incoherent term. Thorndahl at CERN realized that an appropriate filter (Fig. 5) added would achieve this end. In this scheme the cooling signal derives from orbital period variations from the mean. If $2T_0 \eta (\overline{P^2})^{1/2} \lesssim \text{B.W.}^{-1}$ (system bandwidth) a linear approximation is still valid. This is the situation for the CERN A.A. ring. The equivalent of Eq. 16 is, in this approximation:

$$S = \sum_{i=1}^N P_i f(t_i) + P f(o) \quad (19)$$

where $f(t)$ is now constrained (Fig. 5) such that $f(o) = 1$ to be consistent with (15). Now, averaging gives:

$$\frac{\overline{\Delta P^2}}{\overline{P^2}} = -2g + g^2 \overline{\sum_{i=1}^N f^2(t_i)} \quad (20)$$

where we can consistently define the coefficient of g^2 as a constant N_s . The conclusion is that the noiseless maximum momentum cooling rate is twice that of comparable transverse cooling (Eq. 7).

The optimum cooler will have a noise B.W. no larger than its system bandwidth. This follows since P.U. B.W. (determining N_s) is

to be minimized consistent with the machine aperture, which determines a similar kicker geometry and thus noise B.W. Therefore, a valid model of electronic noise in the system (this assumption is equivalent to that represented in Eq. 12) is a train of random pulses identical with the beam particle pulses. In this form the noise behaves as an addition to the incoherent term which I write as rP (r being the sum over noise pulses).

In the average (Eq. 20) only the $g^2 r^2 P^2$ term survives, and the model implies that $r^2 = N_n$, the number of noise pulses contributing to the single passage kick. The noisy result is:

$$\frac{\overline{\Delta P^2}}{P^2} = -(2g - (N_s + N_n)g^2) \quad (21)$$

The filter not only damps the incoherent signal but also the noise in the same manner. For momentum cooling noise contributes to the maximal cooling rate ($g_{\max} = (N_s + N_n)^{-1}$) rather than to the asymptotic beam size.

Now I derive the CERN AA design performance based on their design report parameters.⁴ N_s derives from N , the ring size and the B.W. of 250 MHz.

$$\begin{aligned} N_s &= \frac{N}{2} (B.W.)^{-1} (T_o)^{-1} = (2.5 \times 10^7) (250 \text{ MHz} \cdot 0.52 \text{ } \mu\text{s})^{-1} \\ &= 9.5 \times 10^4 \end{aligned} \quad (22)$$

Since N_s and N_n are proportional respectively to the electronic output power due to Schottky signal and to amplifier noise we get,

$$N_n/N_s = 3.2 \text{ KW}/1.8 \text{ KW}$$

again from CERN results. Now, using Eq. 21 with g set maximally:

$$\tau_{e2} = 2 \cdot 2 \cdot N_s \left(1 + \frac{3.2}{1.8}\right) T_o = 1.1s \quad (23)$$

where e^2 is approximately the cooling ratio anticipated in the design during 2s. Although close to optimum ($N_n \lesssim N_s$) the AA design is still electronic noise limited ($N_n \approx 2N_s$). Since a rather conservative amplifier noise figure is chosen for the design it is likely that this factor two can be reduced.

The discrepancy between (23) and the expected 2s cooling time comes from the fact that the AA design does not approach good mixing. The condition for 50% mixing is:

$$1/2 (B.W.)^{-1} = 2 (\Delta P/P)_{full} \cdot \eta \cdot T_o$$

whereas the RHS ($1.5 \times 10^{-2} \cdot 1/12 \cdot 5 \times 10^{-7}s$) is ~
LHS ($2 \times 10^{-9}s$).

IV. Optimization

A. transverse cooling

The similarity of (21) and (7) implies that the result (8) ought to scale to (23). In fact, (8) scales ($N \rightarrow \frac{1}{40} N$; $L \rightarrow 3.3 L$) to 0.95s (compared to 1.1s). Equation (7) seems to imply an intrinsic factor two slower rate for transverse compared (Eq. (21)) to momentum cooling. In the analysis of Section I only the $\cos\psi_i$ projection of the beam signal was utilized. There is another and entirely statistically independent prejection ($S \propto \sum \sin\psi_i$). Thus we can always cool twice in one revolution with two P.U./K. pairs $\pi/2$ out of phase (for instance, the 2nd P.U. could be at the 1st K. position except for the crosstalk problem). Taking this into account the noiseless transverse and filter-momentum cooling processes have identical results.

Is there an advantage in having two "orthogonal" P.U./K. pairs acting on the beam over the "multiple" P.U. scheme of the AA. (see Fig. 6)? The answer is yes since for orthogonal P.U.'s the incoherent term is also reduced in proportion. For identical electronic channels and for fixed correction per turn $\overline{A^2}(\infty)/\overline{A^2}(0)$ is reduced by $2(1-\sqrt{2})^2 \approx 1/3$ whereas two ganged P.U.'s reduce the same ratio by 1/2. Besides this temporal orthogonality W. Hardt of CERN has noted that a complete set of spatial independent P.U.'s is possible.⁵ Since only the quadrupole P.U. is truly orthogonal to simple dipole plates let me consider an arrangement with just these elements ($\pi/2$ betatron space pairs each). Figure 7 describes the performance of such a system. Another assumption used, as carried over from Eq. (8), is that all channels have a 1 GHz B.W. which would seem to be about the limit of feasibility.

For example, cooling the initial size by $1/e^2$ in 3s requires each of the four channels to have $\sim 5^{\circ}\text{K}$ noise temperature. Solid state preamplifiers could probably be obtained with 100°K noise temperature (LN_2 cooled). This implies ~ 20 ganged P.U./preamp. inputs for each of the four channels.

B. Mixing and bandwidths

If any relative phase mixing occurs between P.U. and K. during transverse cooling, the correction signal (2) is degraded. A careful study of the timing involved in the filter momentum cooling reveals that P.U.-K. longitudinal mixing has the same deliterious effect. (This may be confusing since for the naïve momentum cooling of Fig. 5 such mixing is necessary. The timing of Fig. 6 assumes no P.U.-K. mixing.) We have a design criteria that P.U./K. pairs be as close as possible. It would be useful, in this context, to design machine lattices where given (P.U.-K.) segments have

$$\langle \alpha_p \rangle_{\text{segment}} = 0.$$

The exact amount of longitudinal mixing (\Leftrightarrow phase mixing in this paper) is determined by the system bandwidth (see Appendix). For initial momentum spread $(\overline{P^2}(0))_{\text{full}}^{1/2}$ we have the criteria:

$$2 \cdot T_0 \cdot \eta \cdot (\overline{P^2}(0))_{\text{full}}^{1/2} \geq (\text{B.W.})^{-1} \quad - \text{ transverse.}$$

$$2 \cdot T_0 \cdot \eta \cdot (\overline{P^2}(0))_{\text{full}}^{1/2} \lesssim (\text{B.W.})^{-1} \quad - \text{ momentum (filter).}$$

Obviously, simultaneous, maximally fast momentum and transverse cooling are incompatable. As the momentum spread shrinks, phase mixing disappears.

Ultimately the (B.W.) is determined by the physical P.U. dimensions. In order to be uniformly sensitive, the P.U. structure must have lateral dimensions about equal to the local machine

aperture. Even if some array of small pad electrodes were devised, the average impact parameter would be of the same size. At B.W. ~ 1 GHz this limits aperture already to 30 cm. For a very large emittance pre cooler, maintaining B.W. must be accomplished with low β points at P.U. and K. However, this is contradictory to the requirement of minimal P.U. \rightarrow K. mixing. Incorporation of special machine distortions - either low β points or arbitrary segments with $\langle \alpha_p \rangle_{\text{seg.}} = 0$ - interferes with placement of large numbers of P.U.'s and K's.

The extremely high gains involved in stochastic cooling necessitate very low leakage of kicker signals down the beam line (toward P.U.). The aperture limitation on B.W. is also the point (in frequency) where the pipe will significantly propagate as a guide. The walls may be made resistive to counter this effect but I believe that doing so will distort the beam particle's fields in just such a way to keep the net signal B.W. unimproved.

C. γ dependence

Is there an energy dependence to stochastic cooling? The general answer is no, none, as long as the particles are relativistic ($\beta \sim 1$). This is remarkable when contrasted to the sharp γ dependence of electron friction cooling.

Since the field pulse produced at a (stationary) point by a passing charge has energy density $\propto \gamma$ and width $\propto \gamma^{-1}$ we might expect better cooling at higher energy from either improved signal to noise or from increased B.W. For any relativistic particle this Weisäker-Williams power spectrum is flat and has a high frequency cut off above the aperture limited P.U./K. system

bandwidth. That is, no γ variation will be reflected in the correction signal. Cooling performance is entirely determined by

- 1) N - total number of particles;
- 2) P.U. and K. aperture and
- 3) preamplifier noise figure.

Appendix

A. Transverse P.U.

Figure 3 represents the transverse P.U. structure and input equivalent circuit. The geometry reflects the constraint of aperture limited performance ($d \sim$ aperture).

$$C_o = \epsilon_o L_{eff} \quad (A1)$$

$$V(A) = \frac{2e}{C_o + C_s} \left(\frac{A}{d} \right)$$

This illustrates the obvious need to minimize parasitic shunt capacitance (large P.U. tanks). If the preamplifier input were equivalent to a large resistance (with the usual Johnson noise) then the P.U. capacitance would have a profound influence on the output noise spectrum (high frequencies shorted out). Actually the preamplifiers of interest (F.E.T.) have noise dominated by source-drain shot statistics - entirely independent of input circuit.⁶ The standard description of this is an additional noise voltage source $\langle V_n \rangle$.

When $d/2 \lesssim L$ and for relativistic charges ($\beta \sim 1$) it is legitimate to approximate the time structure of $V(A)$ as a square pulse:

Thus within this model's applicability $(B.W.)^{-1} = 2(\text{transit time})$. Preamps are conveniently and standardly rated by noise temperature. This means

$$\text{R.M.S. power out due to noise} = h^2 4KT (B.W.)$$

which defines T. We want the value of V_n , so assuming a 50 Ω output impedance:

$$\langle V_n^2 \rangle = 4KT \text{ (B.W.) } 50 \Omega \quad (A2)$$

which is indistinguishable from a Schottky signal:

$$\langle V(A)^2 \rangle = \frac{2e^2}{C_o^2} \frac{N_s A^2}{d^2}$$

Thus,

$$r^2 = \frac{d^2 C_o^2}{e^2} \text{ KT (B.W.) } 50 \Omega \quad (A3)$$

The above lumped analysis assumes that the preamp is electrically within $\sim \frac{C}{2(B.W.)}$ of the P.U., which is reasonable for (B.W.) $\lesssim 1$ GHz. If an intervening transmission line does exist, its only effect (assuming it is matched to the preamp input) will be to degrade the signal due to its mismatch to the P.U. The transmission line itself, although it is a real impedance, produces no Johnson noise. The simplicity of the P.U. circuit and neglect of C_s is equivalent to assuming a perhaps unrealistically high effective input impedance for the preamplifier, viz 377 Ω as is emphasized by the form of equation 14.

B. Momentum Signal

Assume Gaussian shape for the pulse produced by each charge passing the P.U. (Fig. 5)

$$\begin{aligned} V_D(t) &= + \exp [-t^2/\sigma^2] && \text{- direct pulse} \\ V_E(t) &= - \exp [-(t+\epsilon)^2/\sigma^2] && \text{- echo pulse} \end{aligned} \quad (A4)$$

where $\epsilon = T_O \cap P$ is the echo delay for the delay filter with delay T_O . The signal to the kicker is:

$$\frac{dV_D(t)}{dt} + \frac{dV_E(t)}{dt} = - \frac{2}{\sigma^2} [t V_D(t) + (T+\epsilon) V_E(t)] \quad (A5)$$

within the linearizing approximation $\delta > \sigma$ we have

$V_D(t) \approx V_E(t) \approx 1$ for $t < \sigma$. Eq. (A4) is then

$$\approx + \frac{4}{\sigma^2} \epsilon$$

which is the correct symmetry linear signal. Notice that (A4) is consistent with zero P.U.-K. dispersion.

Also within the above assumptions it can be seen that a uniform (with P_i) form of $f(t)$ in Fig. 5 is justified since:

$$\frac{\partial}{\partial t} [V_D(t) - V_E(t)] = 0$$

has the solutions (zero crossing points)

$$t = \frac{-\epsilon \pm (\epsilon^2 + 2\sigma^2)^{1/2}}{2}$$

so that the width is approximately constant

$$\sigma_K^2 \equiv \sigma^2 + \epsilon^2/2 .$$

V. Acknowledgements

The idea of Stochastic Cooling originated with Van der Meer⁶ and has been further developed since then at CERN. My work has profited extensively from discussions with F. Mills and P. McIntyre. A. Ruggiero contributed a seminal note on the subject.⁷

References

1. Stochastic Damping of Betatron Oscillations in the ISR, S. Van der Meer, CERN/ISR-PO/72-31.
2. Cern Courier 18, 111 (1978).
3. Design Study of a \bar{P} -P Colliding Beam Facility, CERN/PS/AA 78-3.
4. G. Carron and L. Thorndahl, CERN/ISR-RF/LT/PS.
5. W. Hardt, CERN/PS/DL 76-10.
6. Low Noise Electronic Design, C.D. Motchenbacher, F.C. Fitchen, and John Wiley, 1973.
7. Stochastic Cooling with Noise, A. Ruggiero, Fermilab Tech. Note, October 3, 1977.

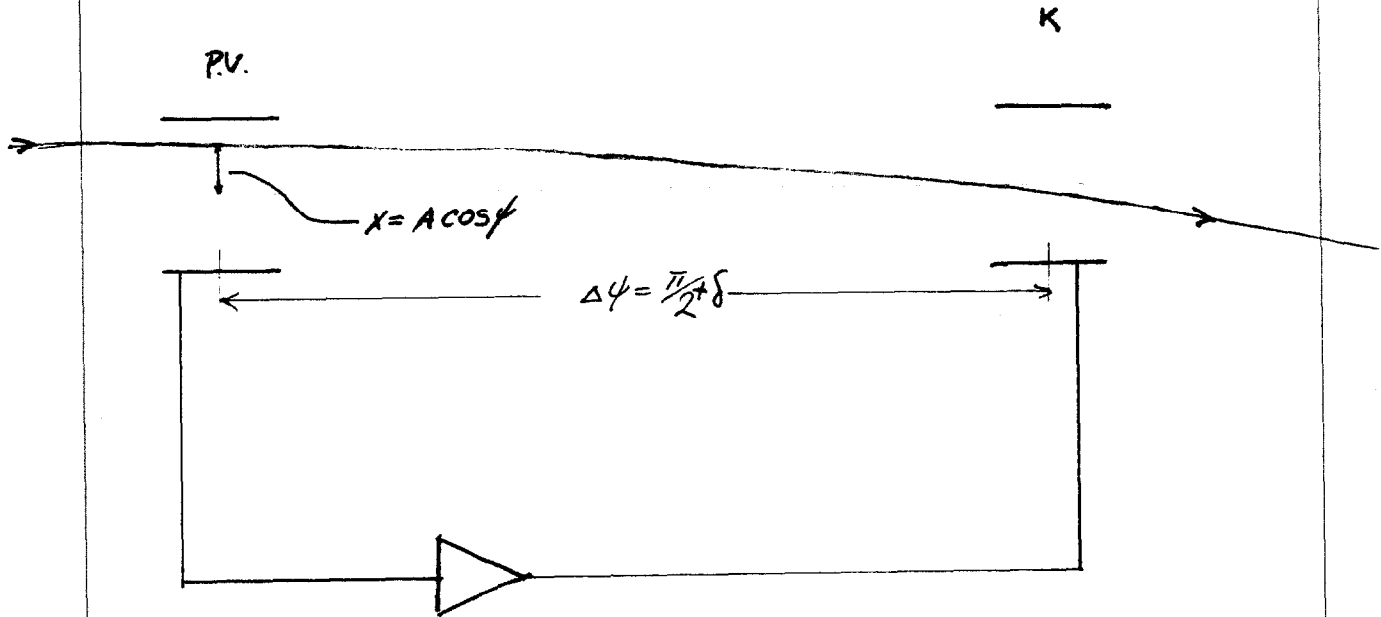
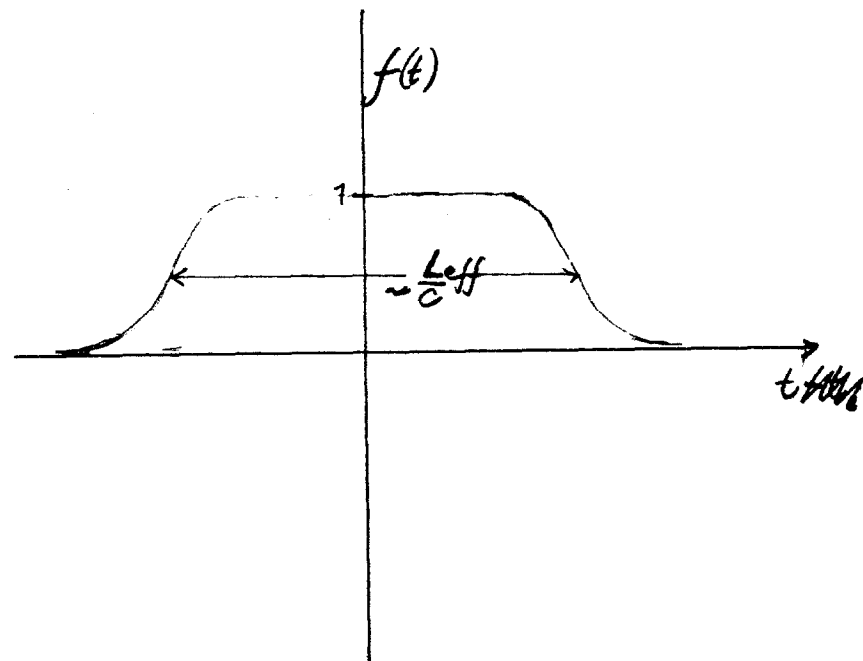
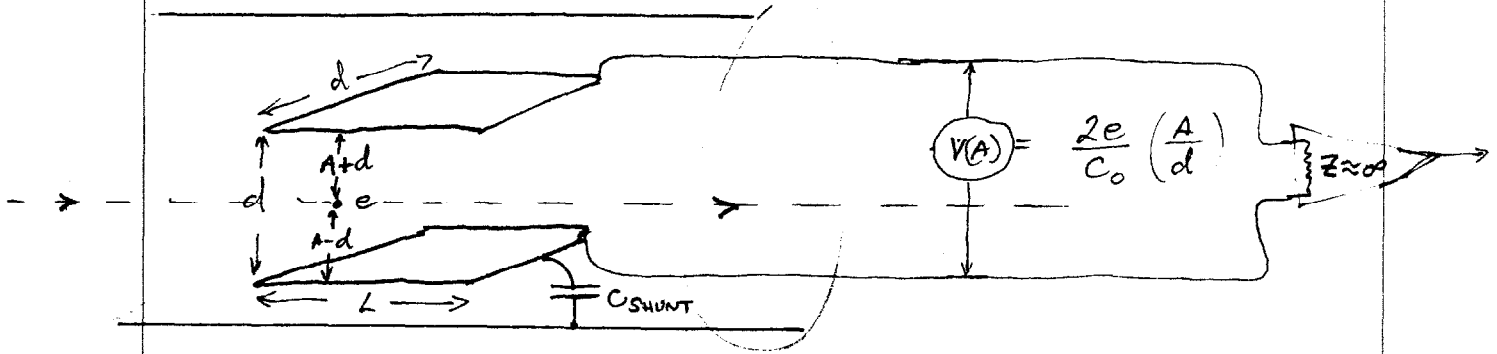


FIG. 2



F163

P.U. CIRCUIT



$$C_0 \equiv \epsilon_0 L_{eff}$$

FIG. 4

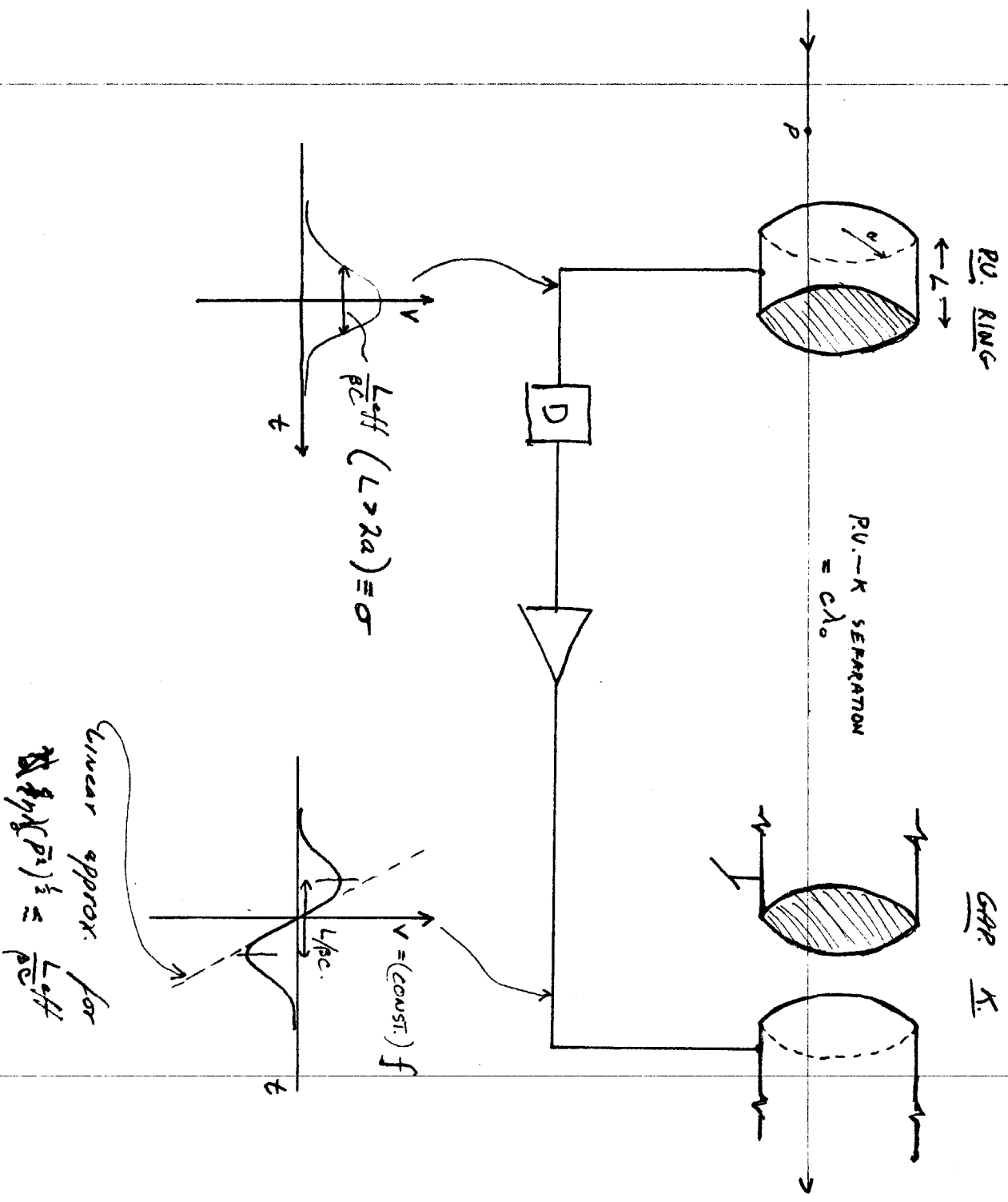
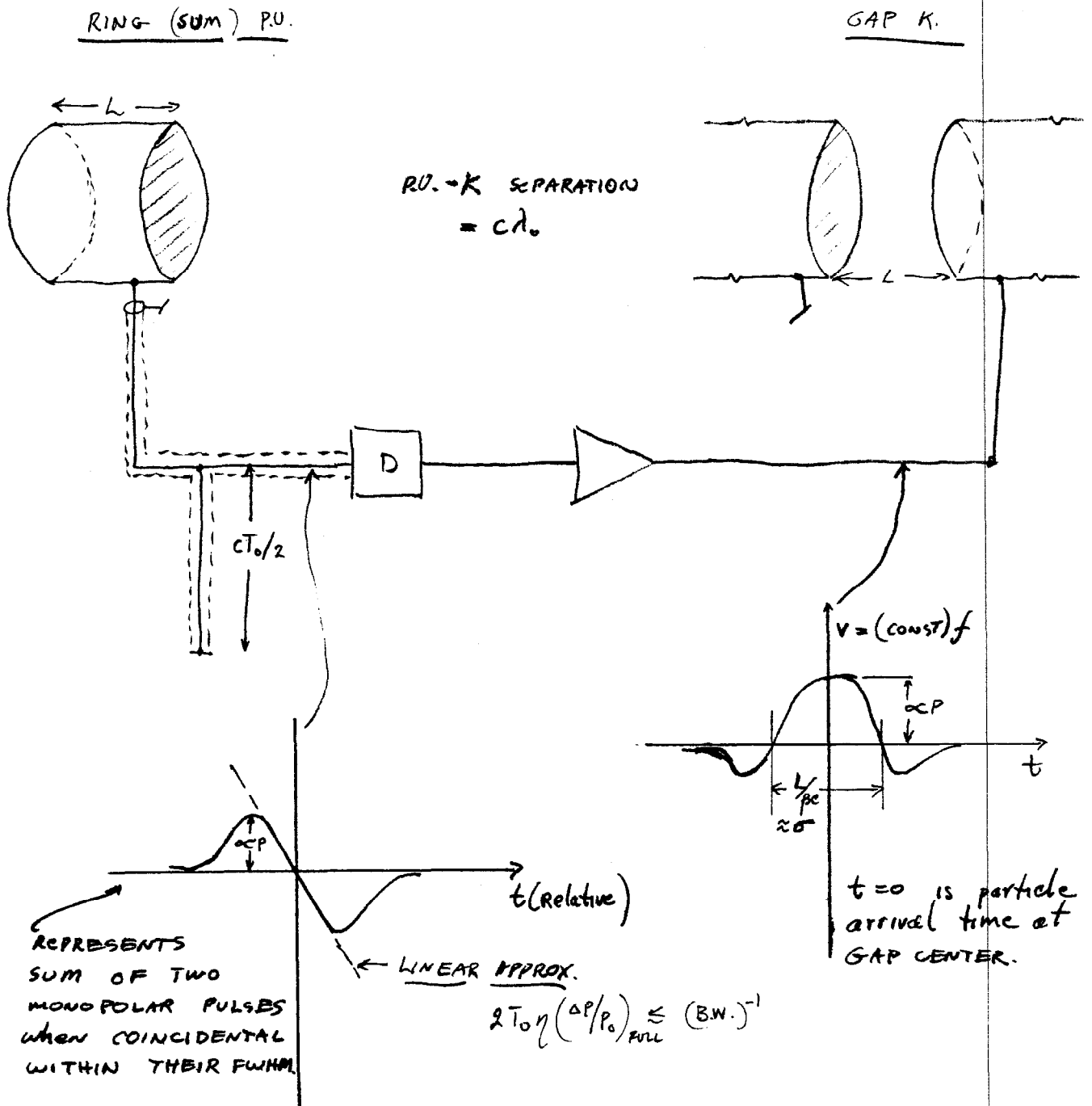
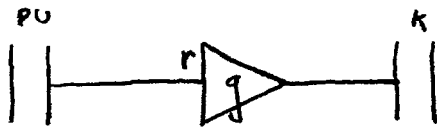


FIG. 5

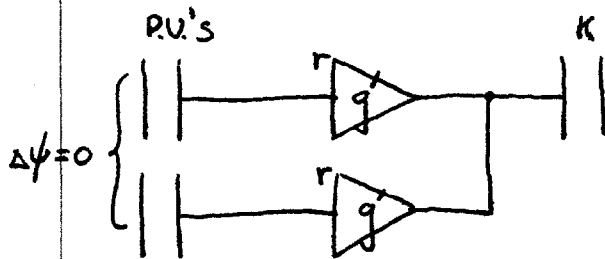


① BASIC CHANNEL



$$\text{CORRECTION} = g^2 r^2 - (g - \frac{1}{2} N_s g^2) \bar{A}^2$$

② MULTIPLE P.V. NOISE REDUCTION

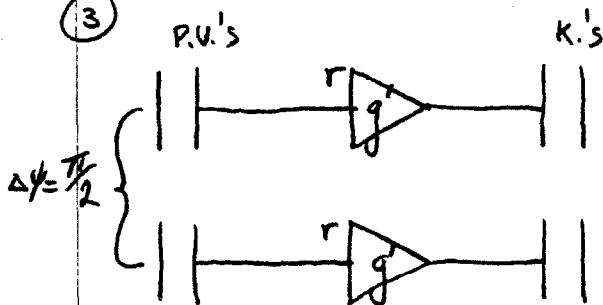


$$\text{CORRECTION} = 2g'^2 r^2 - (2g' - 2N_s g'^2) \bar{A}^2$$

$$g' = \frac{1}{2} g \Rightarrow \text{same RATE} \Rightarrow$$

$$\text{CORRECTION} = \boxed{\frac{1}{2} g^2 r^2 - (g - \frac{1}{2} N_s g^2) \bar{A}^2}$$

③

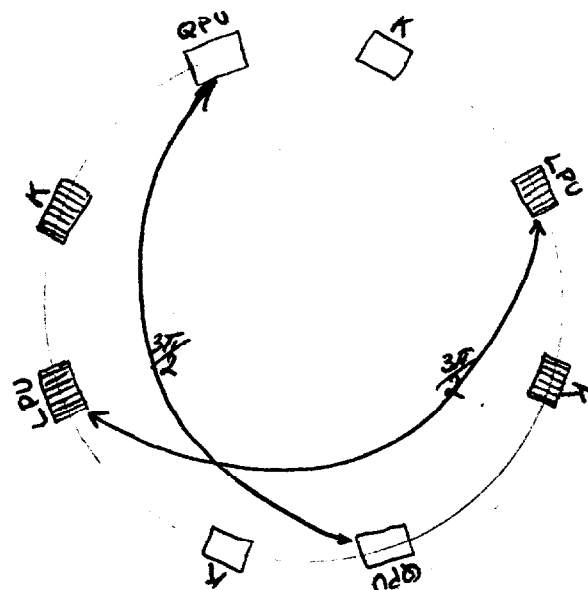


$$\text{CORRECTION} = 2 \left[g'^2 r^2 - (g' - \frac{1}{2} N_s g'^2) \bar{A}^2 \right]$$

$$g' = 1 - \sqrt{2} g \Rightarrow \text{same RATE} \Rightarrow$$

$$\text{CORRECTION} = \boxed{2(1 - \sqrt{2})^2 g^2 r^2 - (g - \frac{1}{2} N_s g^2) \bar{A}^2}$$

FIGURE 7



LPU - LINEAR RF.

QPU - QUADROPOLE RF.

